JOURNAL OF MULTIVARIATE ANALYSIS 11, 400–417 (1981)

Simultaneous Estimation of Poisson Means

JAMES H. ALBERT

Department of Mathematics and Statistics,
Bowling Green State University, Bowling Green, Ohio 43403

Communicated by A. Cohen

In the simultaneous estimation of means from independent Poisson distributions, an estimator is developed which incorporates a prior mean and variance for each Poisson mean estimated. This estimator possesses substantially smaller risk than the usual estimator in a region of the parameter space and seems superior to other estimators proposed to estimate p Poisson means. It is indicated through two asymptotic results that, unlike the conjugate Bayes estimator, the risk of the estimator does not greatly exceed the risk of the usual estimator outside of the region of risk improvement.

1. INTRODUCTION

In this paper the simultaneous estimation of means from independent Poisson distributions is considered. Assume $X_1,..., X_p$ are independent and X_i is distributed Poisson with mean λ_i, $i = 1,..., p$. It is desired to estimate $\lambda = (\lambda_1,..., \lambda_p)$ using an estimator $\delta = (\delta_1,..., \delta_p)$. The loss $L_1(\delta, \lambda) = \sum_{i=1}^{p} (\delta_i - \lambda_i)^2$ will usually be considered. Letting $X = (X_1,..., X_p)$, the usual estimator of λ is $\delta^0(X) = X$, which is the maximum likelihood estimator (MLE) and the minimum variance unbiased estimator.

It is well known that the MLE is inadmissible in estimating a multivariate normal mean (of dimension at least 3) under squared error loss. Similarly, in the Poisson estimation problem, Peng (1979) and Hudson (1978) each showed δ^0 is inadmissible (under loss L_1) for $p \geqslant 3$. Peng's estimator, shown to possess uniformly smaller risk than δ^0, is defined componentwise as

$$\delta_i^P(X) = X_i - \frac{(p - N_0 - 2)_+}{S} b_{X_i}, \qquad i = 1,..., p,$$

where $N_k =$ number of $\{X_j: X_j = k\}$, $b_{X_i} = \sum_{j=1}^{X_i} j^{-1}$, $S = \sum_{i=1}^{p} b_{X_i}^2$, and $(a)_+ = \max\{0, a\}$. The estimators proposed by Hudson and Peng shrink X towards the origin and therefore show most of their risk improvement over δ^0

Received January 8, 1980; revised January 8, 1981.
AMS 1980 subject classifications: Primary 62C25; Secondary 62F15.
Key words and phrases: Improvement region; inadmissible; proportional risk; simultaneous estimation; Stein-type estimator.

0047-259X/81/030400–18$02.00/0

near the origin. Tsui (1978) extended Peng's and Hudson's results by finding an estimator which uniformly improves upon δ^0 and shrinks X towards a positive integer K. His estimator is defined componentwise as

$$\delta_i^T(X) = X_i - \frac{(p - \sum_{n=0}^{K} N_n - 2)_+}{S^*} b_{X_i}^*, \qquad i = 1,\ldots, p,$$

where $b_{X_i}^*$ is a term depending on K and $S^* = \sum_{i=1}^{p} b_{X_i}^{*2}$.

Using the loss function $L_2(\delta, \lambda) = \sum_{i=1}^{p} \lambda_i^{-1}(\delta_i - \lambda_i)^2$, Clevenson and Zidek (1975) developed an estimator with uniformly smaller risk than δ^0 for $p \geqslant 2$. Their estimator is defined componentwise as

$$\delta_i^Z(X) = \left(1 - \frac{\gamma + p - 1}{\sum_{j=1}^{p} X_j + \gamma + p - 1}\right) X_i, \qquad i = 1,\ldots, p,$$

where $1 \leqslant \gamma \leqslant p - 1$. Note that, like δ^P, δ^Z shrinks X towards the origin. If $\gamma > 1$, they showed that δ^Z is Bayes with respect to some proper prior and admissible. Ghosh and Parsian (1980) generalized Clevenson and Zidek's result by obtaining a larger class of proper Bayes estimators dominating δ^0 for $p \geqslant 3$. Recently, Tsui (1980) constructed a class of estimators with uniformly smaller risk than δ^0 using the more general loss function $L_2^c(\delta, \lambda) = \sum_{i=1}^{p} c_i \lambda_i^{-1}(\delta_i - \lambda_i)^2$, where $c_i > 0$, $i = 1,\ldots, p$.

Our aim is to find the most appropriate alternative to δ^0 in the simultaneous estimation of p Poisson means. The estimator δ^0 is minimax under loss L_2, so we cannot expect to find an estimator which substantially improves upon it (with respect to risk) over the entire parameter space. Most alternative estimators, such as those described above, shrink X towards a point and show substantial improvement only in a region about that point. Therefore if a user wants to find a good alternative to δ^0, he should be able to specify the region in which he would like the substantial improvement to occur. In other words, the input of prior information seems necessary in the development of good alternative estimators. This rationale for inputting prior information in improved estimators is discussed in detail in Berger (1980).

In Section 2, an alternative estimator to δ^0 is developed which incorporates a prior mean μ_i and a prior variance $\mu_i \beta_i$ for the component λ_i, $i = 1,\ldots, p$. This estimator is defined componentwise as

$$\delta_i^*(X) = \mu_i + (1 - c^*(x)/(\beta_i + 1))(x_i - \mu_i), \qquad i = 1,\ldots, p,$$

where

$$c^*(x) = \min \left\{1, \frac{\sum_{j=1}^{p} X_j/(\beta_j + 1)}{\sum_{j=1}^{p} X_j/(\beta_j + 1)^2 + \sum_{j=1}^{p} ((X_j - \mu_j)/(\beta_j + 1))^2}\right\}$$

In Section 3, δ^* is shown to be a very attractive alternative to δ^0 when prior information is available. Unlike the estimators δ^P, δ^T, and δ^Z, δ^* appears to simultaneously incorporate prior information and improve significantly over δ^0 in a region of the parameter space. Also, in two asymptotic results, δ^* is shown to be a preferable estimator to the conjugate Bayes estimator when the prior information has been misspecified and the true parameter value lies far from the prior mean. Although δ^* does not possess uniformly smaller risk than δ^0, it is shown that the risk of δ^* does not greatly exceed the risk of δ^0 outside of the improvement region.

In Section 4, we discuss how to use the estimator δ^*. In particular, we discuss how to obtain the set of prior parameters $\{(\mu_i, \beta_i),\ i = 1,..., p\}$. Section 5 illustrates the use of δ^* in an example of estimating the mean numbers of fires during different days of the week simultaneously in New York City.

2. Development of the Estimator

To develop an estimator which possesses substantially smaller risk than δ^0 in a region of the parameter space, first consider the Bayes estimator of λ using the conjugate prior. Assume a priori that $\lambda_1,..., \lambda_p$ are independent with λ_i having the gamma distribution with parameters α_i and β_i. That is, assume $\lambda_1,..., \lambda_p$ have the prior density

$$g(\lambda_1,..., \lambda_p) = \prod_{i=1}^{p} \frac{e^{-\lambda_i/\beta_i}\lambda_i^{\alpha_i - 1}}{\beta_i^{\alpha_i}\Gamma(\alpha_i)}, \quad \lambda_1,..., \lambda_p > 0,\ \alpha_i, \beta_i > 0,\ i = 1,..., p.$$

Under loss L_1, the Bayes estimator of λ_i is

$$\delta_i^B(X) = \frac{(X_i + \alpha_i)\beta_i}{\beta_i + 1} = \alpha_i\beta_i + \left(1 - \frac{1}{\beta_i + 1}\right)(X_i - \alpha_i\beta_i).$$

Note that δ_i^B shrinks the observation X_i towards the prior mean $\alpha_i\beta_i$. An estimator is desired which shrinks towards the prior mean like δ^B, but restricts the amount of shrinkage when the observations appear to contradict the prior information. An estimator with this property should intuitively have smaller risk than the risk of δ^0 in a particular region of the parameter space, and risk not much larger than the risk of δ^0 outside of the region. We, therefore, consider estimators of the form

$$\delta_i(X) = \mu_i + \left(1 - \frac{c(X)}{\beta_i + 1}\right)(X_i - \mu_i), \qquad i = 1,..., p,$$

where $\mu_i = \alpha_i\beta_i$ and $c(X)$ is a function of $X_1,..., X_p$.

To find an appropriate $c(X)$, an argument similar to one in Hudson (1974) is used. If $c(X)$ is temporarily assumed to be a constant c, the risk of δ under loss L_1 can be evaluated to be

$$R(\delta, \lambda) = \sum_{i=1}^{p} \left(1 - \frac{c}{\beta_i + 1}\right)^2 \lambda_i + c^2 \sum_{i=1}^{p} \left(\frac{\lambda_i - \mu_i}{\beta_i + 1}\right)^2.$$

Minimizing the above expression with respect to c shows that the optimal c is

$$c' = \frac{\sum_{i=1}^{p} \lambda_i/(\beta_i + 1)}{\sum_{i=1}^{p} \lambda_i/(\beta_i + 1)^2 + \sum_{i=1}^{p}((\lambda_i - \mu_i)/(\beta_i + 1))^2}.$$

Although c' is a function of the unknown parameters $\lambda_1,..., \lambda_p$, λ_i can be estimated by its MLE X_i, obtaining the estimator

$$\begin{aligned}\delta_i'(X) = \mu_i + &\left(1 - \frac{1}{\beta_i + 1}\right.\\ &\times \left.\frac{\sum_{j=1}^{p} X_j/(\beta_j + 1)}{\sum_{j=1}^{p} X_j/(\beta_j + 1)^2 + \sum_{j=1}^{p}((X_j - \mu_j)/(\beta_j + 1))^2}\right)\\ &\times (X_i - \mu_i), \qquad i = 1,..., p.\end{aligned}$$

Using this method, Hudson derives an estimator similar to δ' with $\beta_1 = \cdots = \beta_p = 0$ and $\mu_1 = \cdots = \mu_p$. The estimator discussed here is an extension of Hudson's estimator, in that it permits a different prior mean and variance input for each of the p Poisson parameters estimated.

Finally, since we would like our estimator to act like a Bayes estimator in a particular region of the parameter space, it is natural to restrict the shrinkage of X_i towards μ_i to the amount that δ^B shrinks X_i. The estimator δ' is then modified to the recommended estimator δ^*.

Consider the behavior of δ^* where the prior information has been misspecified. In this case at least one observation X_i is likely to be far from its prior mean μ_i and $\sum_{j=1}^{p}((X_j - \mu_j)/(\beta_j + 1))^2$ will tend to be large. In this case, $\delta_i^*(X) \cong X_i$, so the estimator is in fact ignoring the wrong prior information. This behavior indicates that δ^* should not have a much larger risk than δ^0 outside the region of improvement.

3. Evaluation

3.1. *Incorporation of Prior Information*

Prior beliefs concerning the set $\{\lambda_1,..., \lambda_p\}$ may be expressed through the parameters $\mu = (\mu_1,..., \mu_p)$ and $\beta = (\beta_1,..., \beta_p)$. In the conjugate Bayesian estimation model described earlier, μ_i and $\mu_i\beta_i$ are respectively the mean and variance of the prior distribution of λ_i. Thus larger values of β_i reflect a flatter and less informative prior distribution.

In the figures that are to be presented, we show that δ^* has a significantly smaller risk than δ^0 in a region about μ and has a risk not much larger than δ^0 elsewhere. Since the risks of δ^*, δ^P, and δ^Z are not expressible in closed form, all of the risks are found numerically using a computer and the values found have a standard error of approximately 5%.

To compare δ^* with δ^0, it is convenient to consider the proportional risk of δ^*, defined by

$$\frac{R(\delta^*,\lambda)}{R(\delta^0,\lambda)}=\frac{E_\lambda^X[\sum_{i=1}^p(\delta_i^*(X)-\lambda_i)^2]}{\sum_{i=1}^p\lambda_i}.$$

Figure 1 presents a graph for $p=2$ showing contours of constant values of proportional risk. Here the prior parameters $(\mu_1,\beta_1)=(\mu_2,\beta_2)=(4,0)$ are used. Keeping in mind that a proportional risk of less than 1 signifies improvement of δ^* over δ^0, one sees that the region of improvement is quite large. From examining δ^* in (1.1), one sees that, by selecting $\beta_1=\beta_2=0$, X is shrunk as far as possible (for our estimators) towards the prior mean (4, 4). If positive values of β_1 and β_2 were used instead in δ^* the proportional risk would increase near the prior mean and the size of the total improvement region would increase. In this example, the proportional risk of

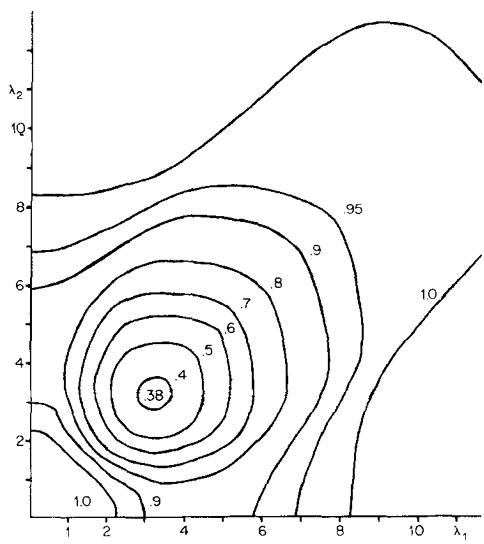

FIG. 1. $p=2$. Contours of constant values of proportional risk of δ^*. Prior information: $(\mu_i,\beta_i)=(4,0)$, $i=1,2$.

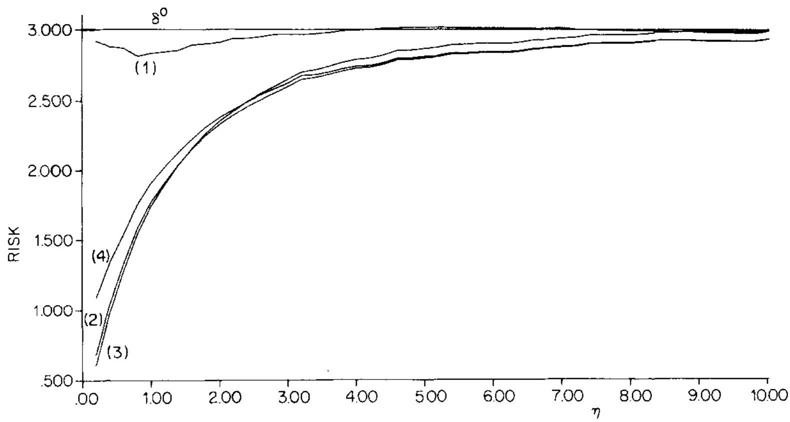

FIG. 2. $p = 3$. Prior information: $\mu_i = \beta_i = 0$ for all i. Loss L_2. Risks of (1) δ^P, (2) δ^Z, $\gamma = 1$, (3) δ^Z, $\gamma = 2$, and (4) δ^* along line $\lambda_1 = \lambda_2 = \lambda_3 = \eta$.

δ^* outside the outer contour of 1 appears to be bounded above by 1.1, while for small (λ_1, λ_2) beyond the inside contour of 1, the proportional risk appears to be bounded by $\lim_{\lambda_1, \lambda_2 \to 0} R(\delta^*, \lambda)/\sum_{i=1}^{2} \lambda_i = 1.27$.

Let us next compare δ^* with two of the previously proposed estimators, δ^P and δ^Z. Consider the case $p = 3$ and since both Peng's and Clevenson and Zidek's estimators were designed to shrink X towards the origin, we initially set $(\mu_i, \beta_i) = (0, 0)$ for each component of δ^*. Figure 2 compares the risk functions of δ^Z with $\gamma = 1$, δ^Z with $\gamma = p - 1 = 2$, and δ^P and δ^* along the diagonal line $\lambda_1 = \lambda_2 = \lambda_3$ for the loss L_2. (This is the loss for which δ^Z was designed.) In this example, the two versions of δ^Z and δ^* appear to be roughly equivalent in terms of risk. One also notes that δ^P only improves marginally on δ^0—it does not shrink X significantly towards the origin. In Albert (1979) a similar risk comparison is made between δ^T and δ^*. In the example considered, δ^T, like δ^P only offers marginal risk improvement over δ^0 near the prior mean, and δ^* has a substantially smaller risk than δ^T near μ.

It is important to note that the estimators of Peng and Clevenson and Zidek have the disadvantage of not being able to accept arbitrary prior means. Tsui's estimator does shrink towards an arbitrary prior mean, but it is not able to improve substantially upon δ^0 in the prior region. The estimator δ^* is able to accept different prior means, and although it does not improve upon δ^0 uniformly, it does improve upon δ^0 significantly in a prior region.

3.2. *Behavior of the Estimator for Large λ*

We now begin the investigation of the risk function of δ^* outside of the region where δ^* shows significant improvement over δ^0. The first situation that will be considered is that in which $\lambda_1,\ldots,\lambda_p$ are large and far outside of the improvement region. It is intuitively clear that δ^*, like the conjugate Bayes estimator δ^B, shrinks towards a prior mean and should show risk improvement over δ^0 about that point. But many standard Bayes estimators (like δ^B) exhibit a very large risk relative to the MLE far away from the prior mean, and, therefore, the risk of δ^* in this region is of interest. In Theorem 1, we consider the case in which the prior means $\mu_1,\ldots,\mu_p$ are fixed and $\lambda_1,\ldots,\lambda_p$ go to infinity along the line described by $\lambda_i = k_i\eta$, $i = 1,\ldots,p$. The asymptotic risk improvement (which could be negative) of δ^* over δ^0 is then given.

THEOREM 1. *Let $\lambda_i = k_i\eta$, $k_i > 0$, $i = 1,\ldots,p$. Then the asymptotic improvement in risk of δ^* over δ^0 as $\eta \to \infty$ is*

$$I = \lim_{\eta\to\infty} [R(\delta^0,\lambda) - R(\delta^*,\lambda)]$$
$$= \frac{[\sum_{j=1}^p k_j/(\beta_j+1)]^2}{\sum_{j=1}^p (k_j/(\beta_j+1))^2} - \frac{4\sum_{j=1}^p k_j/(\beta_j+1)\sum_{j=1}^p (k_j/(\beta_j+1))^3}{[\sum_{j=1}^p (k_j/(\beta_j+1))^2]^2} + 2.$$

Proof. See Appendix.

Consider the asymptotic risk improvement of δ^* over δ^0 given in Theorem 1. The risk improvement has been shown to be of the order of a constant for large $\lambda_1,\ldots,\lambda_p$, while the risk of δ^0 is $\sum_{j=1}^p \lambda_j$. Thus the risk improvement is insignificant compared to the risk of δ^0 for large $\lambda_1,\ldots,\lambda_p$. Note next that when $k_1/(\beta_1+1) = \cdots = k_p/(\beta_p+1)$, it can be calculated that $I = p-2$. Thus when many means are estimated simultaneously, δ^* can display smaller risk than δ^0 even for λ far from the prior mean.

In this asymptotic setting, it is of interest to investigate how poorly δ^* can perform relative to δ^0. Let $b_i = k_i/(\beta_i+1)$ for $i = 1,\ldots,p$ and note without loss of generality that one can take $\sum_{j=1}^p b_j^2 = 1$. Then the asymptotic improvement in risk is

$$I = -4\sum_{j=1}^p b_j^3 \sum_{j=1}^p b_j + \left(\sum_{j=1}^p b_j\right)^2 + 2 \geqslant -4\sum_{j=1}^p b_j + \left(\sum_{j=1}^p b_j\right)^2 + 2,$$

the last step following from the fact that $\sum_{j=1}^p b_j^2 = 1$ implies that $\sum_{j=1}^p b_j^3 \leqslant 1$. Now the last expression is a quadratic in $\sum_{j=1}^p b_j$ and achieves a minimum value of -2 at $\sum_{j=1}^p b_j = 2$. Thus $I \geqslant -2$, meaning that δ^* cannot lose more than two units of risk asymptotically compared to δ^0. Through computer simulation studies, it appears that δ^* will do worst asymptotically along lines very close to the edges of the parameter space. Along these lines, all but one of $\lambda_1,\ldots,\lambda_p$ are close to zero, and δ^* is performing much like a one dimensional estimator.

Let us compare the risk of δ^* with the risk of δ^B in this situation. The risk of δ^B can be calculated to be

$$R(\delta^B, \lambda) = \sum_{i=1}^{p} \left(\frac{\beta_i}{\beta_i + 1}\right)^2 \lambda_i + \sum_{i=1}^{p} \left(\frac{\lambda_i - \mu_i}{\beta_i + 1}\right)^2.$$

Observe that the dominant term in $R(\delta^B, \lambda)$ for large λ is $\sum_{i=1}^p ((\lambda_i - \mu_i)/(\beta_i + 1))^2$, which increases quadratically as a function of λ_i, while $R(\delta^0, \lambda) = \sum_{i=1}^p \lambda_i$ increases linearly as a function of λ_i. Hence outside of a particular "prior region" of the parameter space about the prior mean, δ^B will possess a substantially larger risk than δ^0, and the risk decrement becomes more severe as the distance from the prior mean increases. Therefore δ^* is a "safer" estimator to use than δ^B when parameter values far from the prior mean are likely.

3.3. *Behavior of the Estimator for Large Prior Means*

A second situation of interest is the effect of large prior means on the risk behavior of δ^*. We are primarily interested in evaluating how poorly δ^* can perform (in terms of risk) relative to δ^0 outside of the improvement region. Consider the proportional improvement in risk of δ^* over δ^0, defined by

$$\rho^* = \frac{R(\delta^0, \lambda) - R(\delta^*, \lambda)}{R(\delta^0, \lambda)}.$$

It is argued in Albert (1979) that ρ^* appears to be minimized when the prior parameters $\beta_1, \ldots, \beta_p$ are all selected equal to zero and the parameter λ lies within a few standard deviations of the prior mean μ.

Therefore in the case where the prior means are going to infinity, let $\lambda_1, \ldots, \lambda_p$ also go to infinity such that $\mu_i - \lambda_i = O(\lambda_i^{1/2})$. Theorem 2 gives the asymptotic value of ρ^* in this limiting situation, when $\lambda_1, \ldots, \lambda_p$ increase to infinity along a line from the origin.

THEOREM 2. *Let*

$$\delta_i^*(X) = X_i - \frac{(X_i - \mu_i) \sum_{j=1}^p X_j}{\sum_{j=1}^p X_j + \sum_{j=1}^p (X_j - \mu_j)^2}, \qquad i = 1, \ldots, p.$$

Let $\lambda_i = k_i \eta$, $i = 1, \ldots, p$, *where* $\sum_{i=1}^p k_i = 1$. *Let* $\theta_i = \lambda_i - \mu_i$, *and assume* $\lim_{\eta \to \infty} \theta_i \eta^{-1/2} = \theta_i^*$, $i = 1, \ldots, p$. *Then asymptotically, as* $\eta \to \infty$

$$\lim \frac{R(\delta^0, \lambda) - R(\delta^*, \lambda)}{R(\delta^0, \lambda)}$$

$$= E\left[\sum_{i=1}^{p} \left(2(W_i - \theta_i^*) \frac{W_i}{1 + \sum_{j=1}^p W_j^2} - \left(\frac{W_i}{1 + \sum_{j=1}^p W_j^2}\right)^2\right)\right], \quad (3.1)$$

where $W_i \sim N(\theta_i^*, k_i)$, $i = 1, \ldots, p$, *and* $W_1, \ldots, W_p$ *are independent.*

Proof. See Appendix.

In applying this theorem, it is useful to consider the normal estimation problem where $W_1,\ldots, W_p$ are independent with $W_i \sim N(\theta_i^*, k_i)$, $i = 1,\ldots, p$, and the vector $(\theta_1^*,\ldots, \theta_p^*)$ is to be estimated. Consider the Stein-type estimator

$$\tilde{\delta}(W) = \left(1 - \frac{1}{1 + \sum_{j=1}^{p} W_j^2}\right) W,$$

where $W = (W_1,\ldots, W_p)$, and assume $\tilde{\delta}$ is to be compared with the estimator $\delta^0(W) = W$ using the loss $\sum_{i=1}^{p} (\delta_i - \theta_i^*)^2$. The right-hand side of (3.1) is the risk improvement of $\tilde{\delta}$ upon δ^0. Thus in this asymptotic situation, the proportional improvement of the Poisson estimator δ^* is equivalent to the improvement of the normal estimator $\tilde{\delta}$.

To find the maximum risk decrement of $\tilde{\delta}$ (or equivalently the maximum proportional risk decrement of δ^*), Albert (1979) shows heuristically that the risk decrement is maximized when $p = 1$. In this case, (3.1) is the improvement of the one dimensional estimator

$$\tilde{\delta}(W) = \left(1 - \frac{1}{1 + W^2}\right) W$$

over the estimator $\delta^0(W) = W$ in the situation where $W \sim N(\theta^*, 1)$. Through numerical work, we found that

$$\max_{\theta^*} [R(\tilde{\delta}, \theta^*) - R(\delta^0, \theta^*)] = 0.27.$$

Relating this to the Poisson estimation problem, this indicates that asymptotically, under the conditions of Theorem 2, the decrement in risk of δ^* can be no larger than 27% of the risk of δ^0 for all values of p.

In this same asymptotic large mean situation, one can easily calculate the proportional risk improvement of δ^B to be

$$\frac{R(\delta^0, \lambda) - R(\delta^B, \lambda)}{R(\delta^0, \lambda)} = 1 - \sum_{i=1}^{p} \left(\frac{\beta_i}{\beta_i + 1}\right)^2 k_i - \sum_{i=1}^{p} \frac{\theta_i^{*2}}{(\beta_i + 1)^2}.$$

This proportional improvement clearly is monotonically decreasing away from the point $(\theta_1^*,\ldots, \theta_p^*) = (0,\ldots, 0)$. Thus δ^B again can possess a much larger risk than δ^0 away from the prior mean, and therefore δ^B is much more sensitive than δ^* to parameter values outside of the improvement region.

4. Using the Estimator

For each Poisson mean estimated, one inputs two prior parameters, μ_i and β_i. The simplest way to obtain these parameters is to guess at a mean and variance for each coordinate of λ. Since the prior mean and variance of λ_i are μ_i and $\mu_i\beta_i$, respectively, these guesses can be used to obtain μ_i and β_i. Unfortunately, although it may be easy to guess at a prior mean, a prior variance is harder to determine. Subtle characteristics of the prior distribution such as the tail may greatly influence the variance, and it is uncommon to have prior information concerning the tail.

It is often easier, therefore, to specify fractiles of the prior distribution of λ_i, or to assign probabilities to particular areas of the parameter space. These assigned probabilities can lead to values of μ_i and β_i. For example, if the prior distribution of λ_i can be thought to be approximately normal in the central region, then μ_i and β_i can be calculated from the endpoints of an interval which is thought to contain a specific proportion of the prior distribution. If (a, b) is thought to be the interval which contains the middle 50% of the distribution of λ_i, then by solving the equations

$$a = \mu_i - 0.675(\mu_i\beta_i)^{1/2},$$
$$b = \mu_i + 0.675(\mu_i\beta_i)^{1/2},$$

one can obtain μ_i and β_i. Note that prior knowledge of a specific form of the prior density or knowledge of the tails of the distribution are not necessary in using δ^*.

A second method of obtaining the parameters μ_i and β_i is to estimate them from past observations. Let $Y_{i1},..., Y_{im}$ be past independent observations, with Y_{ij} distributed Poisson with mean λ_{ij}, $j = 1,..., m$. Assume it is thought that λ_{ij}, $j = 1,..., m$, are from a common prior distribution with mean μ_i and variance $\mu_i\beta_i$. Then marginally, $Y_{i1},..., Y_{im}$ are independent with common mean μ_i and common variance $\mu_i(\beta_i + 1)$. By using method of moments, reasonable estimates of μ_i and β_i are, therefore, $\hat{\mu}_i = \bar{Y}_i$ and $\hat{\beta}_i = S_i^2/\bar{Y}_i - 1$, where $\bar{Y}_i = m^{-1}\sum_{j=1}^m Y_{ij}$ and $S_i^2 = m^{-1}\sum_{j=1}^m (Y_{ij} - \bar{Y}_i)^2$. (If $\bar{Y}_i > S_i^2$, set $\hat{\beta}_i = 0$.) Note that this empirical Bayes technique was not used in Section 2 in estimating $(\lambda_1,..., \lambda_p)$ since we were uncertain that $\lambda_1,..., \lambda_p$ came from a common prior.

5. Example

Let us assume that a New York City fireman in 1920 is interested in estimating simultaneously the mean numbers of fires for the 7 days of the week in May. That is, he is interested in estimating $\lambda = (\lambda_1,..., \lambda_7)$, where λ_1 = mean number of fires on Sunday in May 1920,..., λ_7 = mean number of fires on Saturday in May 1920. For his data, the fireman observes the number of fires for each day in April. Let $X_{11},..., X_{1n_1}$ denote the numbers of fires for the Sundays in April,..., and $X_{71},..., X_{7n_7}$ denote the numbers of fires for the Saturdays in April. It is reasonable to assume that $X_{i1},..., X_{in_i}$ are independent observations from a Poisson distribution with mean λ_i, $i = 1,..., 7$, and that also the sets $\{X_{ij}, j = 1,..., n_i\}$, $i = 1,..., 7$ are independent.

To estimate λ, one need only consider the sufficient statistics $Z_1 = \sum_{j=1}^{n_1} X_{1j},..., Z_7 = \sum_{j=1}^{n_7} X_{7j}$. Then the maximum likelihood estimator of λ is defined componentwise by $\delta_i^0(Z) = Z_i/n_i$, $i = 1,..., 7$.

Next, assume that in addition to the data in April, the fireman would like to use the numbers of fires for the days in March to estimate λ. Since he suspects that numbers of fires are not homogeneous over the 3 months, he wants to treat the data in March separately from the data in April. One means of accomplishing this is to regard the data in March as prior information and combine the data in April with this prior information to estimate λ. Specifically, he can obtain the prior parameters μ and β from the data in March (using the method described in Section 4) and then calculate δ^* using the data in April.

In calculating δ^*, note first that $Z_1,..., Z_7$ are independently Poisson with means $n_1\lambda_1,..., n_7\lambda_7$, respectively. To estimate the vector $\lambda^* = (n_1\lambda_1,..., n_7\lambda_7)$, observe that the prior mean and prior variance of $n_k\lambda_k$ are $E[n_k\lambda_k] = n_k\mu_k$ and $\mathrm{Var}[n_k\lambda_k] = (n_k\mu_k)(n_k\beta_k)$, and δ^* in this situation is defined componentwise by

$$\delta_i^*(Z) = n_i\mu_i + \left(1 - \frac{1}{n_i\beta_i + 1} g(Z)\right)(Z_i - n_i\mu_i), \qquad i = 1,..., 7,$$

where

$$g(Z) = \min\left\{1, \frac{\sum_{j=1}^7 Z_j/(n_j\beta_j + 1)}{\sum_{j=1}^7 Z_j/(n_j\beta_j + 1)^2 + \sum_{j=1}^7((Z_j - n_j\mu_j)/(n_j\beta_j + 1))^2}\right\}.$$

A reasonable estimator of $\lambda = (\lambda_1,..., \lambda_7)$ is then $\delta^{**} = (n_1^{-1}\delta_1^*,..., n_7^{-1}\delta_7^*)$. When $n_1 = \cdots = n_7$, it is easy to show that the risk performance of δ^{**} (relative to the maximum likelihood estimator) under the loss $\sum_{i=1}^p (\delta_i - \lambda_i)^2$ is equivalent to the risk performance of δ^* under the loss $\sum_{i=1}^p (\delta_i - n_i\lambda_i)^2$. In this example, the n_i's are nearly equal and therefore the discussion in Sections 2 and 3 is relevant to the performance of the estimator δ^{**}.

Using the numbers of fires in March and April (as given in *The New York Times*), we have calculated the estimators δ^0 and δ^{**} and displayed their values in Table I. This table also gives the values of the prior parameters which were calculated using the March data. Observe that for six of the seven components, the value of δ^{**} is closer than δ^0 to the value of λ.

Finally, we have repeated the above procedure and estimated simultaneously the mean numbers of fires for the 7 days of the week for every month in 1920. In each case the numbers of fires in the previous month were observed and the numbers of fires in the second previous month

TABLE I

Prior Information, Computed Estimators, and Parameter Values for Fire Example where May is the Predicted Month

	Sun	Mon	Tue	Wed	Thu	Fri	Sat
μ	20.25	17.80	17.20	23.40	19.75	19.25	14.25
β	0.81	0	0	0	1.97	0	0
δ^0	15.75	16.75	17.50	13.25	19.40	20.80	21.75
δ^{**}	15.89	16.89	17.46	14.61	19.40	20.59	20.74
λ	17.00	17.60	13.50	19.50	15.00	16.25	15.40

TABLE II

Sums of Squared Errors for Maximum Likelihood Estimator and Estimator δ^{**} for 12 Fire Estimation Problems

Prior month	Observed month	Predicted month	$\sum(\delta^0-\lambda)^2$	$\sum(\delta^{**}-\lambda)^2$	Ratio
Nov	Dec	Jan	226.93	230.85	1.02
Dec	Jan	Feb	449.39	415.14	.92
Jan	Feb	Mar	64.65	61.89	.96
Feb	Mar	Apr	183.24	157.01	.86
Mar	Apr	May	137.73	108.04	.78
Apr	May	Jun	101.32	84.65	.84
May	Jun	Jul	94.18	71.08	.76
Jun	Jul	Aug	83.05	61.92	.75
Jul	Aug	Sep	64.43	44.86	.70
Aug	Sep	Oct	29.12	14.00	.48
Sep	Oct	Nov	373.14	339.86	.91
Oct	Nov	Dec	267.76	251.81	.94

were used as the prior information. Table II gives in each case the sums of the squared errors of the estimators δ^{**} and δ^0 in estimating λ and the ratios of the two sums. Note that the sum of squared errors of δ^* is smaller than that of δ^0 in eleven of the twelve estimation problems, and the ratio of sums of squared errors is as low as 0.48 (when October is the predicted month). This example illustrates how δ^{**} can incorporate vague prior information and perform better than the maximum likelihood estimator δ^0.

Appendix

Outlines of the proofs of Theorems 1 and 2 are given below. The detailed proofs are given in Albert (1979).

Proof of Theorem 1. Without loss of generality, we can set $\mu_1 = \cdots = \mu_p = 0$ and ignore the truncation in the shrinking constant of δ^*. Then δ^* is defined by

$$\begin{aligned} \delta_i^*(X) &= X_i - \phi_i(X) && \text{if not all } X_i = 0 \\ &= 0 && \text{if } X_1 = \cdots = X_p = 0,\ i = 1,\ldots, p, \end{aligned}$$

where

$$\phi_i(X) = \frac{(X_i/(\beta_i+1)) \sum_{j=1}^p X_j/(\beta_j+1)}{\sum_{j=1}^p X_j/(\beta_j+1)^2 + \sum_{j=1}^p (X_j/(\beta_j+1))^2}.$$

Clearly

$$\begin{aligned} I &= R(\delta^0, \lambda) - R(\delta^*, \lambda) \\ &= \sum_{i=1}^p E[2(X_i - \lambda_i)\,\phi_i(X) - (\phi_i(X))^2] \\ &= \sum_{\mathscr{X}} \sum_{i=1}^p [2(x_i - k_i\eta)\,\phi_i(x) - (\phi_i(x))^2] \prod_{j=1}^p \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!} \\ &= I_V + I_{V^c}, \end{aligned}$$

where I_V is the sum over the region $V = \{x: |x_i - k_i\eta| > \eta^{9/16}$ for at least one $i\}$, and I_{V^c} is the sum over V^c. It can be shown that $|\phi_i(x)| \leqslant K$, where K is some constant. Hence

$$I_V \leqslant K_1 \sum_V \sum_{i=1}^p |x_i - k_i\eta| \prod_{j=1}^p \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!} + K_2 \sum_V \prod_{j=1}^p \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!}$$

for constants K_1 and K_2. Now say $|x_m - k_m\eta| > \eta^{9/16}$. Then for large enough chosen integer l,

$$\begin{aligned} I_V &\leqslant K_1 \sum_{\mathscr{X}} \sum_{i=1}^p |x_i - k_i\eta|\,|x_m - k_m\eta|^{l_\eta - 9l/16} \prod_{j=1}^p \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!} \\ &\quad + K_2 \sum_{\mathscr{X}} |x_m - k_m\eta|^l \eta^{-9l/16} \prod_{j=1}^p \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!} \\ &= o(1) \end{aligned} \tag{A.1}$$

(since $E\,|X_m - k_m\eta|^c = O(\eta^{c/2})$ for nonnegative integer c).

Next consider I_{V^c}, and note that $V^c = \{x: |x_i - k_i\eta| \leqslant \eta^{9/16}$ for all $i\}$. Through a succession of Taylor's expansions for $2(x_i - k_i\eta)\,\phi_i(x) - (\phi_i(x))^2$, one can show

$$\begin{aligned} I_{Vc} &= \sum_{Vc} \sum_{i=1}^{p} [2(x_i - k_i\eta)\,\phi_i(x) - (\phi_i(x))^2] \prod_{j=1}^{p} \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!} \\ &= \sum_{Vc} \left\{ \frac{2}{\sum_{j=1}^{p} (k_j/(\beta_j+1))^2} \right. \\ &\quad \times \left[\sum_{j=1}^{p} k_j/(\beta_j+1) \sum_{j=1}^{p} k_j(x_j - k_j\eta)/(\beta_j+1) \right. \\ &\quad + \frac{1}{\eta} \sum_{j=1}^{p} k_j/(\beta_j+1) \sum_{j=1}^{p} (x_j - k_j\eta)^2/(\beta_j+1) \\ &\quad \left. + \frac{1}{\eta} \sum_{j=1}^{p} k_j(x_j - k_j\eta)/(\beta_j+1) \sum_{j=1}^{p} (x_j - k_j\eta)/(\beta_j+1) \right] \\ &\quad - \frac{4}{\eta[\sum_{j=1}^{p} (k_j/(\beta_j+1))^2]^2} \sum_{j=1}^{p} k_j/(\beta_j+1) \\ &\quad \times \sum_{j=1}^{p} k_j(x_j - k_j\eta)/(\beta_j+1) \sum_{j=1}^{p} k_j(x_j - k_j\eta)/(\beta_j+1)^2 \\ &\quad \left. - \frac{(\sum_{j=1}^{p} k_j/(\beta_j+1))^2}{\sum_{j=1}^{p} (k_j/(\beta_j+1))^2} \right\} \prod_{j=1}^{p} \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!} + o(1). \end{aligned} \tag{A.2}$$

Write $I_{Vc} = I_1 - I_2$, where

$$I_1 = \sum_{\mathscr{X}} \{\ \} \prod_{j=1}^{p} \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!}, \qquad I_2 = \sum_{V} \{\ \} \prod_{j=1}^{p} \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!},$$

and $\{\ \}$ denotes the expression in brackets in (A.2). Using Chebychev arguments as in (A.1), it can be shown that

$$I_2 = o(1). \tag{A.3}$$

Next, using the independence of $X_1,\ldots, X_p$ and the facts $E(X_j - k_j\eta) = 0$, $E(X_j - k_j\eta)^2 = k_j\eta$, $j = 1,\ldots, p$, it is easy to show that

$$\begin{aligned} I_1 &= \frac{2[(\sum_{j=1}^{p} k_j/(\beta_j+1))^2 + \sum_{j=1}^{p} (k_j/(\beta_j+1))^2]}{\sum_{j=1}^{p} (k_j/(\beta_j+1))^2} \\ &\quad - \frac{4\sum_{j=1}^{p} k_j/(\beta_j+1) \sum_{j=1}^{p} (k_j/(\beta_j+1))^3}{[\sum_{j=1}^{p} (k_j/(\beta_j+1))^2]^2} - \frac{(\sum_{j=1}^{p} k_j/(\beta_j+1))^2}{\sum_{j=1}^{p} (k_j/(\beta_j+1))^2} \\ &= \frac{(\sum_{j=1}^{p} k_j/(\beta_j+1))^2}{\sum_{j=1}^{p} (k_j/(\beta_j+1))^2} - \frac{4\sum_{j=1}^{p} k_j/(\beta_j+1) \sum_{j=1}^{p} (k_j/(\beta_j+1))^3}{[\sum_{j=1}^{p} (k_j/(\beta_j+1))^2]^2} + 2. \end{aligned} \tag{A.4}$$

Combining (A.1), (A.3), and (A.4) gives the desired result.

Proof of Theorem 2. Let

$$\phi_i(X) = \frac{(X_i - \mu_i) \sum_{j=1}^p X_j}{\sum_{j=1}^p X_j + \sum_{j=1}^p (X_j - \mu_j)^2}, \qquad i = 1,..., p.$$

Note that $R(\delta^0, \lambda) = \sum_{i=1}^p \lambda_i = \eta \sum_{i=1}^p k_i = \eta$. Thus

$$\begin{aligned} I &= \frac{R(\delta^0, \lambda) - R(\delta^*, \lambda)}{R(\delta^0, \lambda)} \\ &= \eta^{-1} \sum_{i=1}^p E[2(X_i - k_i\eta)\, \phi_i(X) - (\phi_i(X))^2] \\ &= \sum_{\mathscr{X}} \sum_{i=1}^p [2(x_i\eta^{-1/2} - k_i\eta^{1/2})\, \phi_i(x)\, \eta^{-1/2} \\ &\quad - (\phi_i(x)\, \eta^{-1/2})^2] \prod_{j=1}^p \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!}. \end{aligned} \tag{A.5}$$

First, since we have assumed that $\lim_{\eta\to\infty} \theta_i\eta^{-1/2} = \theta_i^* = O(1)$ for all i, there exist constants K_1 and $N > 1$ such that $|\theta_i|\, \eta^{-1/2} < K_1$ for all i when $\eta \geqslant N$. Define the constant K by $K = 3K_1$ and write $I = I_V + I_{V^c}$, where I_V is the sum over the region $V = \{x: |x_i - k_i\eta| > K\eta^{9/16}$ for at least one $i\}$ and I_{V^c} is the sum over V^c.

Consider the case where $x \in V$. In this region, one can show for some constant K_2, that $|\phi_i(x)\eta^{-1/2}| \leqslant K_2$ for all i, and therefore for a sufficiently large integer l

$$\begin{aligned} I_V &\leqslant \Bigg| \sum_V \sum_{i=1}^p [2(x_i\eta^{-1/2} - k_i\eta^{1/2})\, \phi_i(x)\, \eta^{-1/2} \\ &\quad - (\phi_i(x)\, \eta^{-1/2})^2] \prod_{j=1}^p \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!} \Bigg| \\ &\leqslant \sum_{\mathscr{X}} \sum_{i=1}^p [2K_2\, |x_i\eta^{-1/2} - k_i\eta^{1/2}| + (K_2)^2]\, |x_m - k_m\eta|^l K^{-l} \eta^{-9/16l} \\ &\quad \cdot \prod_{j=1}^p \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!} \\ &= o(1). \end{aligned} \tag{A.6}$$

Next consider the sum I_{V^c}, and note that $V^c = \{x: |x_i - k_i\eta| \leqslant K\eta^{9/16}$ for all $i\}$. Expanding $\phi_i(x)\eta^{-1/2}$ in a Taylor's series for $x \in V^c$ gives

$$\phi_i(x)\eta^{-1/2} = \phi_i^*(x) + O(\eta^{-7/16}), \tag{A.7}$$

where

$$\phi_i^*(x) = \frac{(x_i - k_i\eta + \theta_i)\eta^{-1/2}}{1 + \sum_{j=1}^p [(x_j - k_j\eta + \theta_j)\eta^{-1/2}]^2}, \qquad i = 1,..., p.$$

Combining (A.6) and (A.7), we have

$$I = \sum_{V^c} \sum_{i=1}^{p} [2(x_i\eta^{-1/2} - k_i\eta^{1/2})\,\phi_i^*(x) - (\phi_i^*(x))^2] \times \prod_{j=1}^{p} \frac{e^{-k_j\eta}(k_j\eta)^{x_j}}{x_j!} + o(1). \tag{A.8}$$

Now we apply the well-known fact (see Makabe and Morimura (1955)) that if $X \sim \text{Poisson}(\lambda)$ with density $p_X(\lambda)$, $y = \lambda^{-1/2}(x - \lambda)$, and $f(y)$ is the density of a standard normal random variable, then when $\lambda \geqslant 1$,

$$p_X(\lambda) = f(y)[\lambda^{-1/2} + \lambda^{-1}(y/2 - y^3/6)] + R, \tag{A.9}$$

where $|R| = L\lambda^{-3/2} + o(\lambda^{-3/2})$ and L is a constant. Using this in (A.8) gives

$$\begin{aligned} I = \sum_{V^c} \sum_{i=1}^{p} &[2(x_i\eta^{-1/2} - k_i\eta^{1/2})\,\phi_i^*(x) - (\phi_i^*(x))^2] \\ &\cdot \prod_{j=1}^{p} ((k_j\eta)^{-1/2} f[(k_j\eta)^{-1/2}(x_j - k_j\eta)] \\ &\times \{1 + (k_j\eta)^{-1/2}[(x_j - k_j\eta)(k_j\eta)^{-1/2}/2 \\ &- (x_j - k_j\eta)^3(k_j\eta)^{-3/2}/6] + R_j\}) + o(1), \end{aligned}$$

where $|R_j| = O(\eta^{-3/2})$, $j = 1,..., p$. It is easy to see that I can be written as $I = I_1 + I_2 + o(1)$, where

$$\begin{aligned} I_1 = \sum_{V^c} \sum_{i=1}^{p} &[2(x_i\eta^{-1/2} - k_i\eta^{1/2})\,\phi_i^*(x) \\ &- (\phi_i^*(x))^2] \prod_{j=1}^{p} (k_j\eta)^{-1/2} f[(k_j\eta)^{-1/2}(x_j - k_j\eta)] \end{aligned}$$

and

$$I_2 = \sum_{V^c} \sum_{i=1}^{p} [2(x_i\eta^{-1/2} - k_i\eta^{1/2})\,\phi_i^*(x) - (\phi_i^*(x))^2] \sum_{l=1}^{3^p-1} \prod_{j=1}^{p} t_{jl},$$

in each product $\prod_{j=1}^{p} t_{jl}$, t_{jl} is either

$$S_j = (k_j\eta)^{-1/2} f[(k_j\eta)^{-1/2}(x_j - k_j\eta)],$$
$$T_j = f[(k_j\eta)^{-1/2}(x_j - k_j\eta)](k_j\eta)^{-1}[(x_j - k_j\eta)(k_j\eta)^{-1/2}/2 - (x_j - k_j\eta)^3(k_j\eta)^{-3/2}/6],$$

or R_j, with at least one of the last two types of terms occurring in each product.

One can show that $|I_2| = o(1)$. Next define $w_i = \eta^{-1/2}(x_i - k_i\eta + \theta_i)$, $i = 1,\ldots, p$, and let $\tilde{\theta}_i = \eta^{-1/2}\theta_i$, $i = 1,\ldots, p$. Then

$$I_1 = \eta^{p/2} \sum_{V^*} \left\{ \sum_{i=1}^{p} \left[2(w_i - \tilde{\theta}_i) \frac{w_i}{1 + \sum_{j=1}^{p} w_j^2} - \left(\frac{w_i}{1 + \sum_{j=1}^{p} w_j^2} \right)^2 \right] \cdot \sum_{j=1}^{p} (2\pi k_j)^{-1/2} \exp\{[(w_j - \tilde{\theta}_j)\, k_j^{-1/2}]^2/2\} \right\}, \tag{A.10}$$

where

$$V^* = \{w\colon w_i = -k_i\eta^{1/2} + \tilde{\theta}_i, -k_i\eta^{1/2} + \tilde{\theta}_i + \eta^{-1/2},\ldots \text{ and } |w_i - \tilde{\theta}_i| \leqslant K\eta^{1/16}\}.$$

Write $I_1 = I_3 - I_4$, where

$$I_3 = \eta^{-p/2} \sum_{A} \{\ \}, \qquad I_4 = \eta^{-p/2} \sum_{A - V^*} \{\ \},$$

$\{\ \}$ is the quantity in brackets in (A.10), and $A = \{w\colon w_i = 0, \pm\eta^{-1/2}, \pm 2\eta^{-1/2},\ldots\}$, and $A - V^* = \{w\colon w_i = 0, \pm\eta^{-1/2}, \pm 2\eta^{-1/2},\ldots\} \cap \{w\colon w_i = -k_i\eta^{-1/2} + \tilde{\theta}_i,\ldots$ for all i, and $|w_j - \tilde{\theta}_j| > K\eta^{1/16}$ for at least one $j\}$. One can show

$$I_4 = o(1) \tag{A.11}$$

and

$$\lim_{\eta\to\infty} I_3 = \lim_{\eta\to\infty} \eta^{-p/2} \sum_{A} \sum_{i=1}^{p} \left[2(w_i - \tilde{\theta}_i) \frac{w_i}{1 + \sum_{j=1}^{p} w_j^2} - \left(\frac{w_i}{1 + \sum_{j=1}^{p} w_j^2} \right)^2 \right] \cdot \prod_{j=1}^{p} (2\pi k_j)^{-1/2} \exp\{[(w_j - \tilde{\theta}_j)\, k_j^{-1/2}]^2/2\}$$
$$= \int_{\mathbb{R}^p} \sum_{i=1}^{p} \left[2(w_i - \theta_i^*) \frac{w_i}{1 + \sum_{j=1}^{p} w_j^2} - \left(\frac{w_i}{1 + \sum_{j=1}^{p} w_j^2} \right)^2 \right] \prod_{j=1}^{p} (2\pi k_j)^{-1/2} \exp\{[(w_j - \theta_j^*)\, k_j^{-1/2}]^2/2\}\, dw_j. \tag{A.12}$$

Combining (A.11) and (A.12) gives the desired result.

ACKNOWLEDGMENTS

The author would like to express his gratitude to Professor James Berger for his constant help and guidance throughout the preparation of this paper. In addition, the author would like to thank the referees for helpful comments.

REFERENCES

[1] ALBERT, J. (1979). *Robust Bayes Estimation*. Mimeograph Series No. 79-9, Department of Statistics, Purdue University.

[2] BERGER, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* **8**, 716–761.

[3] CLEVENSON, M. L., AND ZIDEK, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws. *J. Amer. Statist. Assoc.* **70**, 698–705.

[4] GHOSH, M. AND PARSIAN, A. (1980). *Bayes Minimax Estimation of Multiple Poisson Parameters*. Technical Report, Department of Statistics, Iowa State University.

[5] HUDSON, H. M. (1974). *Empirical Bayes Estimation*. Technical Report No. 58, Department of Statistics, Stanford University.

[6] HUDSON, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6**, 473–484.

[7] PENG, J. C. (1979). Simultaneous estimation of the parameters of independent Poisson distributions. *Ann. Statist.*, in press.

[8] TSUI, K. W. (1978). *Multiparameter Estimation of Discrete Exponential Distributions*. Technical Report No. 37, University of California, Riverside.

[9] TSUI, K. W. (1980). *A Class of Minimax Estimators of Several Poisson Parameters*. Technical Report, Department of Mathematics, The University of British Columbia, Vancouver.

[10] MAKABE AND MORIMA (1955). A normal approximation to the Poisson distribution. Reports on Statistical Application Research, *Juse*, **4**, 37–46.